

Introduction aux méthodes ABC et applications

Julien Stoehr¹

Travail en collaboration avec Pierre Pudlo¹ et Lionel Cuccia¹.

¹I3M, Université Montpellier 2

Rencontres Doctorales Lebesgue 2014

Objectif des méthodes ABC

ABC = approximate Bayesian computation

Objectif

Faire de l'inférence sur les paramètres de modèles complexes.

Paradigme Bayésien

- ▶ Prior $\pi(\theta)$
- ▶ Observation(s) \mathbf{y}_{obs}
- ▶ Vraisemblance ou expression du modèle $f(\mathbf{y}|\theta)$ (incalculable)
- ▶ Posterior

$$\pi(\theta|\mathbf{y}_{\text{obs}}) \propto \pi(\theta)f(\mathbf{y}_{\text{obs}}|\theta)$$

Un algorithme d'acceptation-rejet

```
for  $k = 1 \rightarrow N$  do  
  Générer  $\theta_i$  suivant la loi a priori  $\pi(\theta)$   
  Générer  $\mathbf{y}_i$  suivant la loi du modèle  $f(\mathbf{y}|\theta_i)$   
  if  $\mathbf{y}_i = \mathbf{y}_{\text{obs}}$  then  
    Conserver  $(\theta_i, \mathbf{y}_i)$   
  end if  
end for
```

Résultat : $\pi(\theta|\mathbf{y} = \mathbf{y}_{\text{obs}})$.

Un algorithme d'acceptation-rejet

```
for  $k = 1 \rightarrow N$  do  
  Générer  $\theta_i$  suivant la loi a priori  $\pi(\theta)$   
  Générer  $\mathbf{y}_i$  suivant la loi du modèle  $f(\mathbf{y}|\theta_i)$   
  if  $S(\mathbf{y}_i) = S(\mathbf{y}_{\text{obs}})$  then  
    Conserver  $(\theta_i, \mathbf{y}_i)$   
  end if  
end for
```

Résultat : $\pi(\theta | S(\mathbf{y}) = S(\mathbf{y}_{\text{obs}}))$.

Un algorithme d'acceptation-rejet

```
for  $k = 1 \rightarrow N$  do  
  Générer  $\theta_i$  suivant la loi a priori  $\pi(\theta)$   
  Générer  $\mathbf{y}_i$  suivant la loi du modèle  $f(\mathbf{y}|\theta_i)$   
  if  $\rho(S(\mathbf{y}_i), S(\mathbf{y}_{\text{obs}})) < \epsilon$  then  
    Conserver  $(\theta_i, \mathbf{y}_i)$   
  end if  
end for
```

Résultat : $\pi(\theta | \rho(S(\mathbf{y}_i), S(\mathbf{y}_{\text{obs}})) < \epsilon)$.

Des approximations en série...

1. Posterior : $\pi(\boldsymbol{\theta} \mid \mathbf{y} = \mathbf{y}_{\text{obs}})$
2. Posterior conditionnellement aux statistiques résumées :
 $\pi(\boldsymbol{\theta} \mid S(\mathbf{y}) = S(\mathbf{y}_{\text{obs}}))$
3. Seuil ϵ : $\pi(\boldsymbol{\theta} \mid \rho(S(\mathbf{y}), S(\mathbf{y}_{\text{obs}})) < \epsilon)$
4. Monte-Carlo : échantillon suivant la distribution précédente

Un exemple : les champs de Gibbs cachés

Choisir entre deux structures de dépendance d'un champ de Markov caché est un problème difficile

Dans cet exemple, on présentera

- ▶ les difficultés liées au problème de choix de modèles dans le cas des champs de Gibbs cachés
- ▶ la construction de **statistiques résumées géométriques** permettant de traiter ce problème.

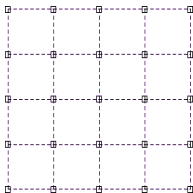
Modèle de Potts caché et choix de modèles

HPM($\mathcal{G}, \alpha, \beta$) ~ modèle de Potts caché

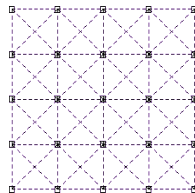
- ▶ \mathcal{G} graph de dépendance,
- ▶ α paramètre de bruit entre le champ caché et le champ observé,
- ▶ β paramètre d'interaction agissant sur les arêtes de \mathcal{G} .

Objectif estimer la probabilité *a posteriori* $\pi(m|y)$

$\mathcal{M}_4 = \text{HPM}(\mathcal{G}_4, \alpha, \beta)$ où \mathcal{G}_4 est



$\mathcal{M}_8 = \text{HPM}(\mathcal{G}_8, \alpha, \beta)$ où \mathcal{G}_8 est



Pour les **champs observés**, c.f. Grelaud, Robert, Marin, Rodolphe and Taly (Bayes. Anal., 2009).

Une vraisemblance à problèmes

La probabilité *a posteriori* d'un modèle est donnée

$$\pi(m|y) \propto \int_{\Theta_m} \frac{1}{Z(\mathcal{G}, \beta)} \frac{1}{C(\alpha)} Q(y, \mathcal{G}, \theta) \pi_m(\theta) d\theta,$$

où

$$C(\alpha) = \{\exp(\alpha) + (K - 1) \exp(-\alpha)\}^N,$$

$$Q(y, \mathcal{G}, \theta) = \sum_{x \in \mathcal{X}} \exp \left\{ \beta \sum_{i \sim j} \mathbb{1}\{x_i = x_j\} + \alpha \left(2 \sum_{i=0}^{N-1} \mathbb{1}\{x_i = y_i\} - N \right) \right\}.$$

Un problème, trois difficultés !

La solution bayésienne au choix de modèles

Paradigme Bayésien

$$\pi(m|y) \propto \int_{\Theta_m} \underbrace{\pi(\mathcal{M} = m) f_m(y|\theta) \pi_m(\theta)}_{(*)} d\theta$$

Methodologie

- ▶ Générer un grand nombre de particules (m, θ_m, y) suivant $(*)$
- ▶ Conserver celles provenant des modèles pour lesquelles $S(y)$ est proche de $S(\mathbf{y}_{\text{obs}})$ au sens d'une distance ρ .

Resultat : un estimateur Monte Carlo standard de $\pi(m | S(\mathbf{y}_{\text{obs}}))$.

$$\hat{\pi}_{ABC}(m | S(\mathbf{y}_{\text{obs}})) = \text{frequence de } m \text{ parmi les particules sélectionnées.}$$

Un équilibre à trouver

$$\pi(m | S(\mathbf{y}_{\text{obs}})) \neq \pi(m | \mathbf{y}_{\text{obs}})$$

Un équilibre est à trouver entre **perte de l'information** et **dimension de $S(\cdot)$** .

- ▶ $S(\cdot)$ de *petite* dimension $\Rightarrow \pi(m | S(\mathbf{y}_{\text{obs}}))$ est une mauvaise approximation.
- ▶ $S(\cdot)$ de *grande* dimension $\Rightarrow \pi(m | S(\mathbf{y}_{\text{obs}}))$ est une bonne approximation mais difficile de générer y tel que $S(y) \approx S(\mathbf{y}_{\text{obs}})$.

Références

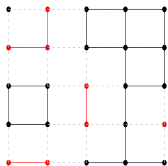
- ▶ Blum (JASA, 2010)
- ▶ Didelot, Everitt, Johansen, and Lawson (Bayes. Anal., 2011)
- ▶ Robert, Cornuet, Marin and Pillai (PNAS, 2011)
- ▶ Fearnhead and Prangle (JRRS-B, 2012)
- ▶ Prangle, Fearnhead, Cox, Biggs and French (SAGMB, 2013)
- ▶ Marin, Pillai, Robert and Rousseau (JRSS-B, 2014)

Quatre statistiques résumées géométriques

$$\Gamma(\mathcal{G}, y) : i \sim j \iff i \stackrel{\mathcal{G}}{\sim} j \text{ et } y_i = y_j$$

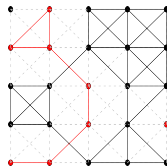
- **nombre de composantes connexes** : $T(\mathcal{G}, y)$
- **taille de la plus grande composante connexe** : $U(\mathcal{G}, y)$

$$\Gamma(\mathcal{G}_4, y)$$



$$T(\mathcal{G}_4, y) = 7$$
$$U(\mathcal{G}_4, y) = 12$$

$$\Gamma(\mathcal{G}_8, y)$$



$$T(\mathcal{G}_8, y) = 4$$
$$U(\mathcal{G}_8, y) = 16$$

Quelles statistiques choisir ?

Objectif : Choisir l'ensemble de statistiques résumées le plus pertinent possible.

Stat. résumées	Grelaud, <i>et al.</i>	Nombre de comp. conn.	Plus grande comp. conn.
dim = 2	✓		
dim = 4	✓	✓	
dim = 6	✓	✓	✓

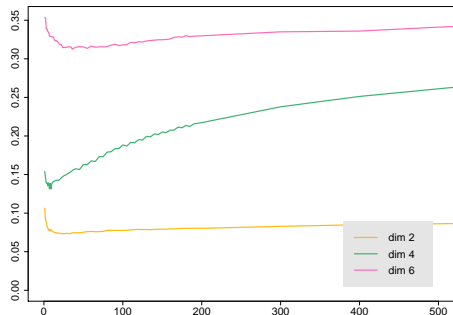
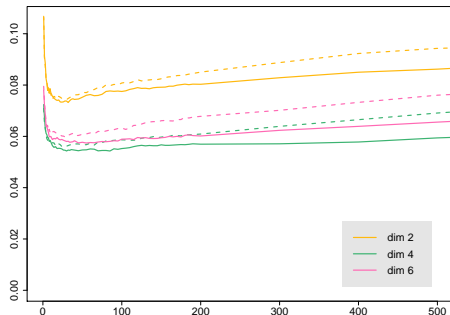
Et quel seuil ϵ ?

Un *bon* point de vue : les k plus proches voisins.

Training : table sur laquelle on entraîne les classifieurs (50 000 ou 100 000).

Validation : table de données dont on connaît l'indice du modèle (20 000).

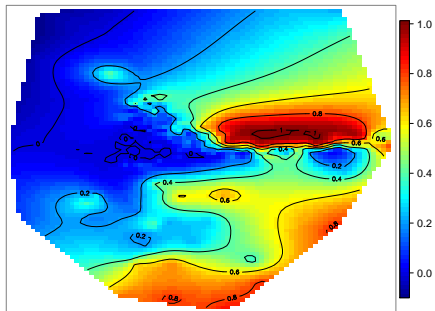
Test : table sur laquelle on teste notre procédure avec le k estimé.



Un taux d'erreur local

TABLE : Evaluation de l'erreur *a priori* sur une table test de taille 30,000.

Train size	5,000	100,000
2D statistics	8.8%	7.9%
4D statistics	6.5%	6.1%
6D statistics	7.1%	7.1%
Adaptive ABC	6.2%	5.5%



Conclusion

- ▶ Calibrer un k plus proche voisin fournit de meilleurs résultats que le choix arbitraire d'un quantile de distance.
- ▶ Une nouvelle classe de statistiques résumant les images à partir de clusters de pixels.
- ▶ Un taux d'erreur local permettant la mise en place d'une procédure adaptative.

Perspectives

- ▶ Etudier plus en détails la construction de statistiques résumées avec des graphes induits notamment lorsque le nombre de couleur augmente.
- ▶ **Publié dans *Statistics and Computing*** *Adaptive ABC model choice and geometric summary statistics for hidden Gibbs random fields* (J. Stoehr, P. Pudlo, L. Cucala)

Merci pour votre attention